Binary Loan Status Classification

Semih Barutcu

2/10/2021

```
library(pacman)
p_load(dplyr, tidyverse, lubridate, Amelia, vtable, tictoc, rpart, rpart.plot, C50, ROCR,
       caret, randomForest, tictoc, ranger, class, gmodels, naivebayes, autoEDA, summarytools)
```

In this project, Lending Club accepted loan data was studied. Loan status was response variable and I worked on both 84 remaining variable and selected 14 remaining variable. Machine Learning algorithms implementations and results can be seen below step by step.

Step 1 – Reading the data

```
tic()
LendingClub  <- read_csv("accepted_2007_to_2018Q4.csv") %>% mutate_if(is.character, as.factor)
toc()
```

```
## 96.88 sec elapsed
```

I eliminated some variables because they are identifier variables train data which possibly overfit. Also, some of variables are not selected because of impractical to use and including excessive NA values.

```
#  year from issue_d and make it integer
LendingClub$year <- str_sub(LendingClub$issue_d, start=-4) %>% as.integer(LendingClub$year)
```

```
## Warning: Unknown or uninitialised column: `year`.
```

```
LendingClub_2012to2014 <- LendingClub %>%
  filter(between(year,2012,2014)) %>%
  select(-id, -member_id, -emp_title, -issue_d, -url, -desc, -zip_code, -title,
         -earliest_cr_line, -last_pymnt_d, -last_credit_pull_d, -total_pymnt, -total_pymnt_inv,
         -total_rec_int)

# Eliminating columns with more than 20% of NAs
LendingClub_2012to2014 <- LendingClub_2012to2014[, colMeans(is.na(LendingClub_2012to2014)) < 0.1]


LendingClub_2012to2014v2 <- LendingClub %>%
  filter(between(year,2012,2014)) %>%
  select(loan_status, funded_amnt, annual_inc, term, installment, int_rate, grade, dti,
         verification_status, fico_range_low, total_acc, tot_cur_bal, acc_open_past_24mths, num_bc_sats

LendingClub_2012to2014v2 <- LendingClub %>%
  filter(between(year,2012,2014)) %>%
  select(loan_status, funded_amnt, annual_inc, term, installment, int_rate, grade, dti,
         verification_status, fico_range_low, total_acc, tot_cur_bal, acc_open_past_24mths, num_bc_sats
         open_acc, revol_bal, revol_util, last_fico_range_low, total_rev_hi_lim,
```

```
        bc_open_to_buy, mo_sin_old_rev_tl_op, mort_acc, num_bc_sats, num_bc_tl, num_sats,
        tot_hi_cred_lim, total_bal_ex_mort)
```

Step 2 – Exploring and preparing the data

From loan_status table, we can see that 3 results were observed at most. I filtered the data just for these options to get more accurate results.

On the first graph, loan status were depicted according to count of funded amounts and faceted by term. 36 months loan users had a right skewed distribution while 60 months users had an uneven distribution.

On the first graph, loan status were depicted according to count of interest rate and faceted by term. 36 months loan users had a right skewed distribution again while 60 months users had normal distribution.

We can see that 60 months term loan users had higher rate of charged off from table and graphs.

```
LendingClub_2012to2014 <- LendingClub_2012to2014 %>%
  filter(loan_status == "Charged Off"  | loan_status == "Fully Paid") %>% na.omit()

LendingClub_2012to2014v2 <- LendingClub_2012to2014v2 %>%
  filter(loan_status == "Charged Off"  | loan_status == "Fully Paid") %>% na.omit()

LendingClub_2012to2014v2$loan_status <- factor(LendingClub_2012to2014v2$loan_status)
levels(LendingClub_2012to2014v2$loan_status)
```

```
## [1] "Charged Off" "Fully Paid"
```

```
addmargins(table(LendingClub_2012to2014v2$loan_status, LendingClub_2012to2014v2$term))
```

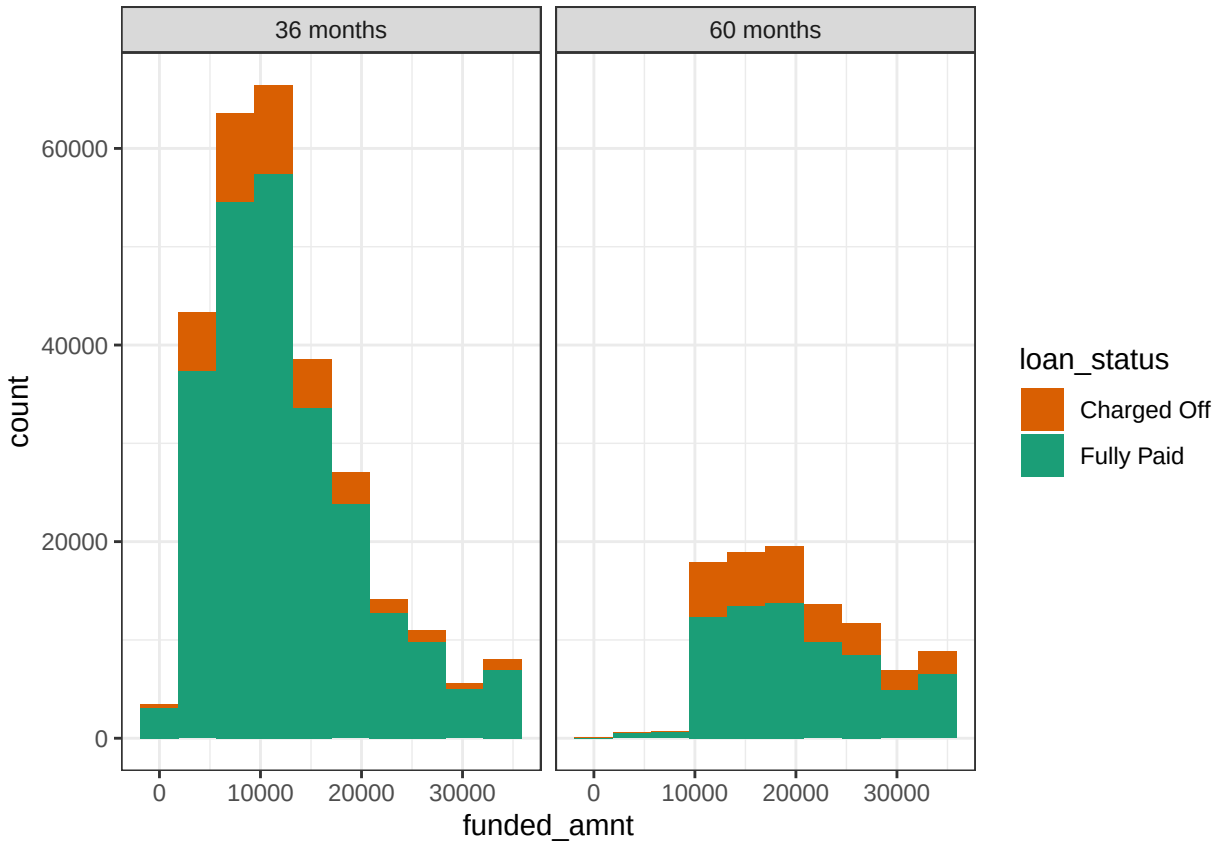
```
##
##               36 months 60 months    Sum
##   Charged Off     36982     28515  65497
##   Fully Paid     244133     70274 314407
##   Sum            281115     98789 379904
```

```
addmargins(prop.table(table(LendingClub_2012to2014v2$loan_status, LendingClub_2012to2014v2$term)))
```

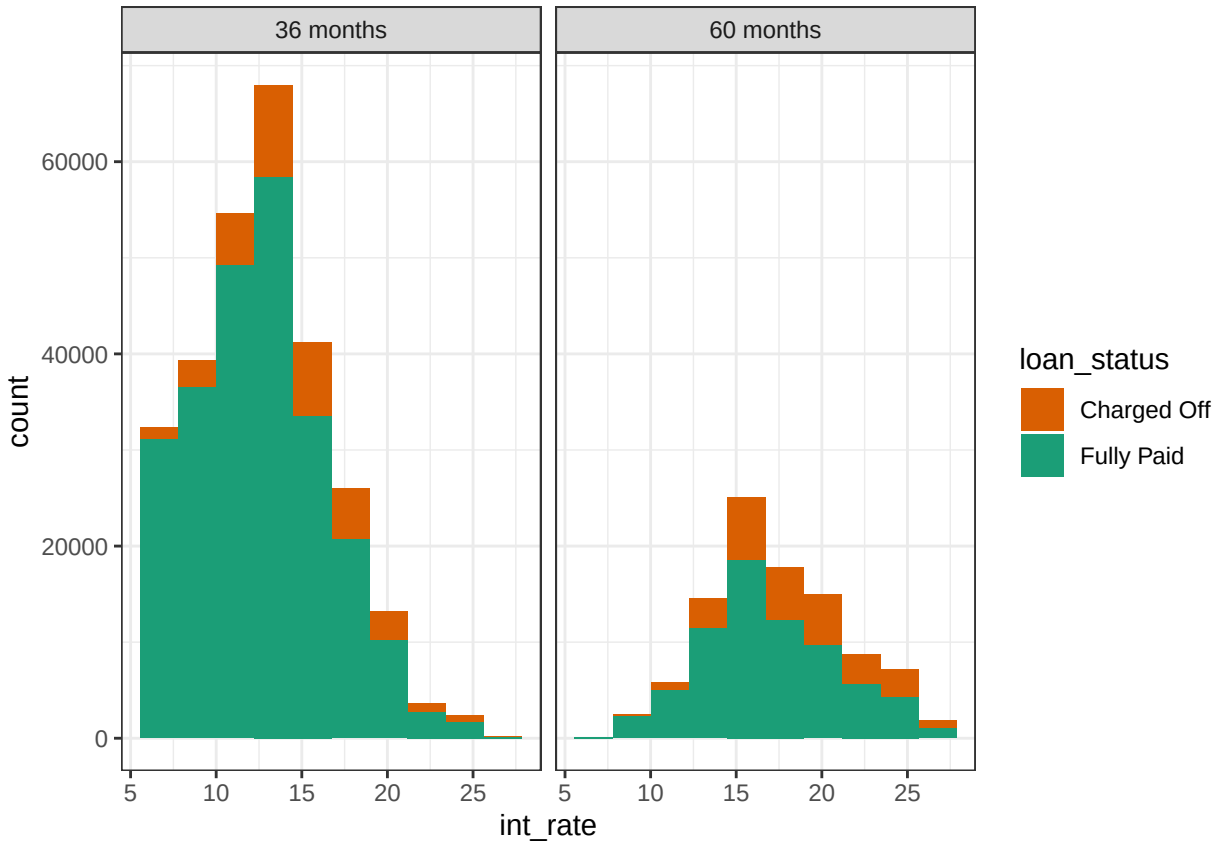
```
##
##                36 months  60 months        Sum
##   Charged Off 0.09734565 0.07505844 0.17240408
##   Fully Paid  0.64261761 0.18497831 0.82759592
##   Sum         0.73996325 0.26003675 1.00000000
```

```
options(repr.plot.width = 1, repr.plot.height = 0.5)

LendingClub_2012to2014v2 %>%
  ggplot(aes(funded_amnt, fill = loan_status)) +
  geom_histogram(bins = 10) +
  scale_fill_brewer(palette = "Dark2", direction = -1) +
  facet_wrap(~term) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```




```
LendingClub_2012to2014v2 %>%
  ggplot(aes(int_rate, fill = loan_status)) +
  geom_histogram(bins = 10) +
  scale_fill_brewer(palette = "Dark2", direction = -1) +
  facet_wrap(~term) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```



Step 3 – Training models

Splitting

The data is imbalanced. Original proportions are 82.7% of fully-paid and 17.3% of charged-off. Good credits observations are under-sampled accordinglt total number of bad credits and a balanced dataset is obtained which has 50% of the each categories. This way, the resulting balanced dataset would provide a better learning process for any model.

The balanced data has 130994 observations and 25 independent variables with the response variable loan_status. I used 75 to 25 percent split for training and test datasets. All of these models are tuned over a validation set sampled within training set without replacement.

```
charged_off <- LendingClub_2012to2014v2 %>% filter(loan_status == "Charged Off")

nrow(charged_off)
```

```
## [1] 65497
```

```
set.seed(123)
fully_paid <- LendingClub_2012to2014v2 %>% filter(loan_status == "Fully Paid")

fully_paid_sample <- sample_n(fully_paid, nrow(charged_off))
```

```
Lending_balanced <- bind_rows(charged_off, fully_paid_sample)
```

```
set.seed(123)
```

```
idx <- sample(nrow(Lending_balanced), round(0.75*nrow(Lending_balanced)))

train_full <- Lending_balanced[idx,]
test <- Lending_balanced[-idx,]

idx2 <- sample(nrow(train_full), round(0.75*nrow(train_full)))

train <- train_full[idx2,]
validation <- train_full[-idx2,]


train_sample <- sample_n(train, 10000)
test_sample <- sample_n(test, 2500)
```

Null Model

```
train %>%
  group_by(loan_status) %>%
  summarise(n = n()) %>%
  mutate(freq = n/sum(n))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 3
##   loan_status     n  freq
##   <fct>       <int> <dbl>
## 1 Charged Off 36843 0.500
## 2 Fully Paid  36841 0.500
```

```
validation %>%
  group_by(loan_status) %>%
  summarise(n = n()) %>%
  mutate(freq = n/sum(n))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 3
##   loan_status     n  freq
##   <fct>       <int> <dbl>
## 1 Charged Off 12306 0.501
## 2 Fully Paid  12256 0.499
```

Logistic Regression

```
log_reg <- glm(loan_status~., train, family = binomial)

summary(log_reg)
```

```
##
## Call:
## glm(formula = loan_status ~ ., family = binomial, data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.0511  -0.6161  -0.0004   0.5748   5.5909
```

```
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -1.138e+01  4.027e-01 -28.256  < 2e-16 ***
## funded_amnt                        2.594e-05  9.649e-06   2.689 0.007177 **
## annual_inc                         2.469e-06  3.358e-07   7.353 1.93e-13 ***
## term60 months                     -8.670e-01  6.181e-02 -14.028  < 2e-16 ***
## installment                       -1.582e-03  2.995e-04  -5.283 1.27e-07 ***
## int_rate                          -2.450e-02  8.891e-03  -2.756 0.005854 **
## gradeB                            -9.077e-02  5.413e-02  -1.677 0.093567 .
## gradeC                            -1.472e-01  7.373e-02  -1.996 0.045934 *
## gradeD                            -1.766e-01  9.628e-02  -1.834 0.066628 .
## gradeE                            -1.787e-01  1.230e-01  -1.453 0.146186
## gradeF                             1.665e-02  1.537e-01   0.108 0.913760
## gradeG                             2.562e-02  1.902e-01   0.135 0.892846
## dti                               -1.824e-02  1.615e-03 -11.291  < 2e-16 ***
## verification_statusSource Verified -4.148e-02  2.815e-02  -1.473 0.140642
## verification_statusVerified        -4.892e-02  2.861e-02  -1.710 0.087243 .
## fico_range_low                    -2.278e-03  5.285e-04  -4.310 1.63e-05 ***
## total_acc                         -2.919e-03  1.827e-03  -1.598 0.110070
## tot_cur_bal                       -7.094e-07  6.016e-07  -1.179 0.238365
## acc_open_past_24mths              -1.711e-02  4.325e-03  -3.957 7.59e-05 ***
## num_bc_sats                        2.937e-02  6.925e-03   4.242 2.22e-05 ***
## open_acc                           5.831e-02  4.480e-02   1.302 0.193069
## revol_bal                          9.277e-07  2.151e-06   0.431 0.666227
## revol_util                         2.608e-03  6.536e-04   3.991 6.59e-05 ***
## last_fico_range_low                2.274e-02  1.748e-04 130.095  < 2e-16 ***
## total_rev_hi_lim                  -2.115e-06  1.831e-06  -1.155 0.248222
## bc_open_to_buy                    -4.528e-06  2.072e-06  -2.186 0.028832 *
## mo_sin_old_rev_tl_op              -6.287e-04  1.277e-04  -4.924 8.47e-07 ***
## mort_acc                           3.079e-03  6.526e-03   0.472 0.637038
## num_bc_tl                         -1.330e-02  3.849e-03  -3.456 0.000548 ***
## num_sats                          -6.677e-02  4.451e-02  -1.500 0.133598
## tot_hi_cred_lim                    1.177e-06  5.623e-07   2.093 0.036375 *
## total_bal_ex_mort                  7.188e-07  3.765e-07   1.909 0.056213 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 102148  on 73683  degrees of freedom
## Residual deviance:  59989  on 73652  degrees of freedom
## AIC: 60053
##
## Number of Fisher Scoring iterations: 6
```

```
pred <- predict(log_reg, newdata = train, type = "response")

log_reg_pred <- ifelse(pred > 0.5, "Fully Paid", "Charged Off")
```

```
CrossTable(x = log_reg_pred, y = train$loan_status, prop.c = F, prop.r = F,
           prop.chisq = FALSE)
```

```
##
##
```

```
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  73684
##
##
##              | train$loan_status
## log_reg_pred | Charged Off |  Fully Paid |   Row Total |
## -------------|-------------|-------------|-------------|
##  Charged Off |       31479 |        5625 |       37104 |
##              |       0.427 |       0.076 |             |
## -------------|-------------|-------------|-------------|
##   Fully Paid |        5364 |       31216 |       36580 |
##              |       0.073 |       0.424 |             |
## -------------|-------------|-------------|-------------|
## Column Total |       36843 |       36841 |       73684 |
## -------------|-------------|-------------|-------------|
##
##
```

```r
mean(train$loan_status == log_reg_pred)
```

```
## [1] 0.8508631
```

```r
pred <- predict(log_reg, newdata = test, type = "response")

log_reg_pred <- ifelse(pred > 0.5, "Fully Paid", "Charged Off")


CrossTable(x = test$loan_status, y = log_reg_pred, prop.c = F, prop.r = F,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  32748
##
##
##                  | log_reg_pred
## test$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------|-------------|-------------|-------------|
##      Charged Off |       13966 |        2382 |       16348 |
##                  |       0.426 |       0.073 |             |
## -----------------|-------------|-------------|-------------|
##       Fully Paid |        2557 |       13843 |       16400 |
```

```
##                  |       0.078 |       0.423 |             |
## -----------------|-------------|-------------|-------------|
##     Column Total |       16523 |       16225 |       32748 |
## -----------------|-------------|-------------|-------------|
##
##
```

```
mean(test$loan_status == log_reg_pred)
```

```
## [1] 0.8491816
```

kNN

kNN method was implemented after required normalization. I chose k as 4 firstly. The accuracies were 85.88% and 77.68% for train and test data respectively. Detailed proportions can be seen on CrossTable. Over-fitting is on the acceptable range for kNN method.

```
# Normalization function
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}

# Prepare data
train_knn <- train %>% mutate_if(is.factor, as.numeric)
test_knn <- test %>% mutate_if(is.factor, as.numeric)

# Normalization
train_knn_n  <- as.data.frame(lapply(train_knn[2:26], normalize))
test_knn_n  <- as.data.frame(lapply(test_knn[2:26], normalize))

# Prediction for train data
pred_knntrain <- knn(train_knn_n, train_knn_n, cl= train_knn$loan_status, k=20)

# Evaluating model performance
CrossTable(x = train_knn$loan_status, y = pred_knntrain,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  73684
##
##
##                       | pred_knntrain
## train_knn$loan_status |         1 |         2 | Row Total |
## ----------------------|-----------|-----------|-----------|
##                     1 |     31429 |      5414 |     36843 |
##                       |     0.853 |     0.147 |     0.500 |
```

```
##                       |     0.819 |     0.153 |           |
##                       |     0.427 |     0.073 |           |
## ----------------------|-----------|-----------|-----------|
##                     2 |      6933 |     29908 |     36841 |
##                       |     0.188 |     0.812 |     0.500 |
##                       |     0.181 |     0.847 |           |
##                       |     0.094 |     0.406 |           |
## ----------------------|-----------|-----------|-----------|
##          Column Total |     38362 |     35322 |     73684 |
##                       |     0.521 |     0.479 |           |
## ----------------------|-----------|-----------|-----------|
##
##
```

```
mean(train_knn$loan_status == pred_knntrain)
```

```
## [1] 0.8324331
```

```
# Prediction for test data
pred_knntest <- knn(train_knn_n, test_knn_n, cl= train_knn$loan_status, k=20)

# Evaluating model performance
CrossTable(x = test_knn$loan_status, y = pred_knntest,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  32748
##
##
##                      | pred_knntest
## test_knn$loan_status |         1 |         2 | Row Total |
## ---------------------|-----------|-----------|-----------|
##                    1 |     13237 |      3111 |     16348 |
##                      |     0.810 |     0.190 |     0.499 |
##                      |     0.788 |     0.195 |           |
##                      |     0.404 |     0.095 |           |
## ---------------------|-----------|-----------|-----------|
##                    2 |      3551 |     12849 |     16400 |
##                      |     0.217 |     0.783 |     0.501 |
##                      |     0.212 |     0.805 |           |
##                      |     0.108 |     0.392 |           |
## ---------------------|-----------|-----------|-----------|
##         Column Total |     16788 |     15960 |     32748 |
##                      |     0.513 |     0.487 |           |
## ---------------------|-----------|-----------|-----------|
```

```
##
##
```

```
mean(test_knn$loan_status == pred_knntest)
```

```
## [1] 0.7965677
```

Boosted C5.0

One of the best way to learn loans is classification trees. Default decision tree via C5.0 can be seen below.

```
train_c50 <- train %>% select(-term, -grade)
validation_c50 <- validation %>% select(-term, -grade)
test_c50 <- test %>% select(-term, -grade)

train_sample_c50 <- train_sample %>% select(-term, -grade)
test_sample_c50 <- test_sample %>% select(-term, -grade)
```

```
modelc50 <- C5.0(loan_status~., train)

modelc50
```

```
##
## Call:
## C5.0.formula(formula = loan_status ~ ., data = train)
##
## Classification Tree
## Number of samples: 73684
## Number of predictors: 25
##
## Tree size: 307
##
## Non-standard options: attempt to group attributes
```

```
summary(modelc50)
```

```
##
## Call:
## C5.0.formula(formula = loan_status ~ ., data = train)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sun Feb 21 00:59:24 2021
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 73684 cases (26 attributes) from undefined.data
##
## Decision tree:
##
## last_fico_range_low > 650:
## :...last_fico_range_low > 685: Fully Paid (22962/841)
## :   last_fico_range_low <= 685:
## :   :...last_fico_range_low <= 665:
## :       :...term = 60 months:
## :       :   :...total_acc <= 15: Fully Paid (121/28)
## :       :   :   total_acc > 15:
```

```
## :       :   :   :...verification_status = Verified:
## :       :   :       :...last_fico_range_low <= 660:
## :       :   :       :   :...total_rev_hi_lim > 64400: Charged Off (47/7)
## :       :   :       :   :   total_rev_hi_lim <= 64400:
## :       :   :       :   :   :...grade in {A,F,G}: Charged Off (48/22)
## :       :   :       :   :       grade = B:
## :       :   :       :   :       :...last_fico_range_low <= 655: Charged Off (11/4)
## :       :   :       :   :       :   last_fico_range_low > 655: Fully Paid (10/3)
## :       :   :       :   :       grade = C:
## :       :   :       :   :       :...fico_range_low <= 680: Fully Paid (34/10)
## :       :   :       :   :       :   fico_range_low > 680: Charged Off (54/15)
## :       :   :       :   :       grade = D:
## :       :   :       :   :       :...num_bc_sats <= 2: Fully Paid (7/1)
## :       :   :       :   :       :   num_bc_sats > 2:
## :       :   :       :   :       :   :...mort_acc <= 7: Charged Off (53/16)
## :       :   :       :   :       :       mort_acc > 7: Fully Paid (3)
## :       :   :       :   :       grade = E:
## :       :   :       :   :       :...int_rate > 21.15: Fully Paid (45/20)
## :       :   :       :   :           int_rate <= 21.15: [S1]
## :       :   :       :   last_fico_range_low > 660:
## :       :   :       :   :...grade in {A,B,G}: Charged Off (19/8)
## :       :   :       :       grade = F:
## :       :   :       :       :...tot_hi_cred_lim <= 79626: Charged Off (6)
## :       :   :       :       :   tot_hi_cred_lim > 79626: Fully Paid (12/5)
## :       :   :       :       grade = D:
## :       :   :       :       :...tot_cur_bal <= 219213: Charged Off (23/6)
## :       :   :       :       :   tot_cur_bal > 219213:
## :       :   :       :       :   :...dti <= 15.46: Charged Off (4/1)
## :       :   :       :       :       dti > 15.46: Fully Paid (10)
## :       :   :       :       grade = C:
## :       :   :       :       :...total_acc > 53: Charged Off (3)
## :       :   :       :       :   total_acc <= 53:
## :       :   :       :       :   :...mort_acc <= 1: [S2]
## :       :   :       :       :       mort_acc > 1:
## :       :   :       :       :       :...open_acc <= 18: Fully Paid (25/3)
## :       :   :       :       :           open_acc > 18: Charged Off (2)
## :       :   :       :       grade = E:
## :       :   :       :       :...open_acc > 17: Charged Off (5)
## :       :   :       :           open_acc <= 17:
## :       :   :       :           :...mort_acc > 2: Charged Off (4/1)
## :       :   :       :               mort_acc <= 2:
## :       :   :       :               :...fico_range_low <= 710: Fully Paid (20/3)
## :       :   :       :                   fico_range_low > 710: Charged Off (2)
## :       :   :       verification_status in {Not Verified,Source Verified}:
## :       :   :       :...last_fico_range_low <= 660:
## :       :   :           :...mo_sin_old_rev_tl_op > 317: Charged Off (34/8)
## :       :   :           :   mo_sin_old_rev_tl_op <= 317:
## :       :   :           :   :...funded_amnt > 30600: [S3]
## :       :   :           :       funded_amnt <= 30600:
## :       :   :           :       :...grade in {A,E}: Charged Off (48/18)
## :       :   :           :           grade in {B,C,G}: Fully Paid (110/46)
## :       :   :           :           grade = D:
## :       :   :           :           :...bc_open_to_buy <= 19858: Fully Paid (67/17)
## :       :   :           :           :   bc_open_to_buy > 19858: Charged Off (9/2)
```

```
## :       :   :           :           grade = F:
## :       :   :           :           :...bc_open_to_buy <= 1113: Fully Paid (10)
## :       :   :           :               bc_open_to_buy > 1113: [S4]
## :       :   :           last_fico_range_low > 660:
## :       :   :           :...grade in {A,D,F}: Fully Paid (59/22)
## :       :   :               grade = G: Charged Off (3/1)
## :       :   :               grade = B:
## :       :   :               :...total_acc <= 33: Fully Paid (5)
## :       :   :               :   total_acc > 33: Charged Off (7/1)
## :       :   :               grade = C:
## :       :   :               :...dti <= 20.81: Fully Paid (33/4)
## :       :   :               :   dti > 20.81: Charged Off (18/8)
## :       :   :               grade = E:
## :       :   :               :...num_bc_sats > 10: Charged Off (2)
## :       :   :                   num_bc_sats <= 10:
## :       :   :                   :...acc_open_past_24mths > 4: Fully Paid (9)
## :       :   :                       acc_open_past_24mths <= 4:
## :       :   :                       :...num_bc_tl <= 9: Charged Off (5)
## :       :   :                           num_bc_tl > 9: [S5]
## :       :   term = 36 months:
## :       :   :...fico_range_low > 715:
## :       :       :...funded_amnt > 15825:
## :       :       :   :...mo_sin_old_rev_tl_op <= 170: Fully Paid (42/16)
## :       :       :   :   mo_sin_old_rev_tl_op > 170:
## :       :       :   :   :...fico_range_low > 725: Charged Off (26/1)
## :       :       :   :       fico_range_low <= 725:
## :       :       :   :       :...dti <= 14.83: Fully Paid (7)
## :       :       :   :           dti > 14.83: Charged Off (16/3)
## :       :       :   funded_amnt <= 15825:
## :       :       :   :...dti > 25.02:
## :       :       :       :...annual_inc <= 41500: Charged Off (10)
## :       :       :       :   annual_inc > 41500: Fully Paid (26/12)
## :       :       :       dti <= 25.02: [S6]
## :       :       fico_range_low <= 715:
## :       :       :...total_acc <= 25: Fully Paid (1260/273)
## :       :           total_acc > 25:
## :       :           :...mo_sin_old_rev_tl_op <= 100: Fully Paid (68/7)
## :       :               mo_sin_old_rev_tl_op > 100:
## :       :               :...last_fico_range_low > 660:
## :       :                   :...grade in {A,B,C,F,G}: Fully Paid (278/65)
## :       :                   :   grade = E:
## :       :                   :   :...total_rev_hi_lim <= 11800: Charged Off (4)
## :       :                   :   :   total_rev_hi_lim > 11800: Fully Paid (5)
## :       :                   :   grade = D:
## :       :                   :   :...verification_status = Not Verified:
## :       :                   :       :...funded_amnt <= 13225: Fully Paid (10)
## :       :                   :       :   funded_amnt > 13225: Charged Off (3)
## :       :                   :       verification_status = Source Verified: [S7]
## :       :                   :       verification_status = Verified:
## :       :                   :       :...revol_util <= 56.1: Charged Off (6)
## :       :                   :           revol_util > 56.1:
## :       :                   :           :...num_bc_sats > 5: Fully Paid (9)
## :       :                   :               num_bc_sats <= 5: [S8]
## :       :                   last_fico_range_low <= 660:
```

```
## :       :                   :...num_bc_tl <= 8: Fully Paid (163/42)
## :       :                       num_bc_tl > 8:
## :       :                       :...grade in {B,D,G}: Fully Paid (264/100)
## :       :                           grade = F: Charged Off (4/1)
## :       :                           grade = A: [S9]
## :       :                           grade = E:
## :       :                           :...tot_cur_bal <= 41353: Fully Paid (8)
## :       :                           :   tot_cur_bal > 41353:
## :       :                           :   :...tot_cur_bal > 237683: Fully Paid (5)
## :       :                           :       tot_cur_bal <= 237683: [S10]
## :       :                           grade = C:
## :       :                           :...int_rate <= 13.66: Fully Paid (39/8)
## :       :                               int_rate > 13.66: [S11]
## :       last_fico_range_low > 665:
## :       :...term = 36 months: Fully Paid (3920/565)
## :           term = 60 months:
## :           :...num_bc_tl > 20:
## :               :...grade in {A,C}: Fully Paid (17/5)
## :               :   grade in {E,G}: Charged Off (8/1)
## :               :   grade = B:
## :               :   :...mo_sin_old_rev_tl_op <= 200: Fully Paid (4)
## :               :   :   mo_sin_old_rev_tl_op > 200: Charged Off (3)
## :               :   grade = F:
## :               :   :...annual_inc <= 47500: Charged Off (2)
## :               :   :   annual_inc > 47500: Fully Paid (2)
## :               :   grade = D:
## :               :   :...last_fico_range_low <= 670: Charged Off (5)
## :               :       last_fico_range_low > 670:
## :               :       :...bc_open_to_buy <= 15331: Fully Paid (5)
## :               :           bc_open_to_buy > 15331:
## :               :           :...revol_bal <= 47231: Charged Off (4)
## :               :               revol_bal > 47231: Fully Paid (2)
## :               num_bc_tl <= 20:
## :               :...fico_range_low > 735:
## :                   :...total_bal_ex_mort <= 61054: Charged Off (49/19)
## :                   :   total_bal_ex_mort > 61054: Fully Paid (24/3)
## :                   fico_range_low <= 735:
## :                   :...open_acc <= 4: Fully Paid (28)
## :                       open_acc > 4:
## :                       :...last_fico_range_low > 675: Fully Paid (661/130)
## :                           last_fico_range_low <= 675:
## :                           :...fico_range_low > 680:
## :                               :...mo_sin_old_rev_tl_op <= 269: Fully Paid (243/77)
## :                               :   mo_sin_old_rev_tl_op > 269:
## :                               :   :...annual_inc <= 170000: Charged Off (58/21)
## :                               :       annual_inc > 170000: Fully Paid (6)
## :                               fico_range_low <= 680:
## :                               :...grade in {A,B,E}: Fully Paid (99/33)
## :                                   grade = F:
## :                                   :...dti <= 29.05: Fully Paid (35/6)
## :                                   :   dti > 29.05: Charged Off (3)
## :                                   grade = G:
## :                                   :...int_rate <= 25.28: Charged Off (2)
## :                                   :   int_rate > 25.28: Fully Paid (12/2)
```

```
## :                                  grade = C:
## :                                  :...open_acc <= 11: Fully Paid (45/1)
## :                                  :   open_acc > 11:
## :                                  :   :...mort_acc <= 2: Fully Paid (17/1)
## :                                  :       mort_acc > 2: [S12]
## :                                  grade = D:
## :                                  :...last_fico_range_low > 670:
## :                                      :...tot_cur_bal > 65499: Fully Paid (33)
## :                                      :   tot_cur_bal <= 65499: [S13]
## :                                      last_fico_range_low <= 670: [S14]
## last_fico_range_low <= 650:
## :...last_fico_range_low <= 610: Charged Off (31582/4315)
##     last_fico_range_low > 610:
##     :...term = 60 months:
##         :...last_fico_range_low <= 640: Charged Off (2757/499)
##         :   last_fico_range_low > 640:
##         :   :...total_acc > 37: Charged Off (190/35)
##         :       total_acc <= 37:
##         :       :...acc_open_past_24mths > 8:
##         :           :...total_rev_hi_lim > 19500: Charged Off (28)
##         :           :   total_rev_hi_lim <= 19500:
##         :           :   :...annual_inc <= 42500: Fully Paid (5)
##         :           :       annual_inc > 42500: Charged Off (11/2)
##         :           acc_open_past_24mths <= 8:
##         :           :...verification_status = Not Verified:
##         :               :...last_fico_range_low <= 645: Charged Off (20/8)
##         :               :   last_fico_range_low > 645: Fully Paid (11/1)
##         :               verification_status = Source Verified:
##         :               :...annual_inc <= 106500: Charged Off (195/69)
##         :               :   annual_inc > 106500: Fully Paid (28/7)
##         :               verification_status = Verified:
##         :               :...last_fico_range_low > 645:
##         :                   :...mort_acc <= 6: Charged Off (164/58)
##         :                   :   mort_acc > 6: Fully Paid (9/2)
##         :                   last_fico_range_low <= 645:
##         :                   :...num_bc_tl > 9: Charged Off (62/6)
##         :                       num_bc_tl <= 9:
##         :                       :...grade = A: Fully Paid (1)
##         :                           grade in {B,C,F,G}: Charged Off (48/9)
##         :                           grade = D:
##         :                           :...tot_cur_bal <= 41864: Charged Off (5)
##         :                           :   tot_cur_bal > 41864: Fully Paid (19/5)
##         :                           grade = E:
##         :                           :...total_acc <= 16: Charged Off (13/2)
##         :                               total_acc > 16: Fully Paid (12/3)
##         term = 36 months:
##         :...last_fico_range_low <= 635:
##             :...total_acc <= 11:
##             :   :...total_rev_hi_lim > 15550:
##             :   :   :...tot_hi_cred_lim <= 282343: Charged Off (93/31)
##             :   :   :   tot_hi_cred_lim > 282343: Fully Paid (5)
##             :   :   total_rev_hi_lim <= 15550:
##             :   :   :...tot_cur_bal > 142209: Fully Paid (20/1)
##             :   :       tot_cur_bal <= 142209:
```

```
##             :   :       :...grade in {A,G}: Fully Paid (8/1)
##             :   :           grade = F: Charged Off (2)
##             :   :           grade = B:
##             :   :           :...open_acc <= 7: Fully Paid (43/17)
##             :   :           :   open_acc > 7: Charged Off (6)
##             :   :           grade = D:
##             :   :           :...tot_cur_bal <= 6637: Fully Paid (10)
##             :   :           :   tot_cur_bal > 6637: Charged Off (41/17)
##             :   :           grade = E:
##             :   :           :...mo_sin_old_rev_tl_op <= 169: Fully Paid (24/3)
##             :   :           :   mo_sin_old_rev_tl_op > 169: Charged Off (3)
##             :   :           grade = C:
##             :   :           :...mort_acc > 0: Fully Paid (12/1)
##             :   :               mort_acc <= 0:
##             :   :               :...last_fico_range_low > 630:
##             :   :                   :...num_bc_tl <= 1: Charged Off (2)
##             :   :                   :   num_bc_tl > 1: Fully Paid (19/2)
##             :   :                   last_fico_range_low <= 630: [S15]
##             :   total_acc > 11:
##             :   :...funded_amnt <= 8525:
##             :       :...num_bc_sats > 6: Charged Off (203/63)
##             :       :   num_bc_sats <= 6:
##             :       :   :...grade in {A,C,G}: Charged Off (455/178)
##             :       :       grade = F:
##             :       :       :...tot_hi_cred_lim <= 33300: Charged Off (6)
##             :       :       :   tot_hi_cred_lim > 33300: Fully Paid (19/6)
##             :       :       grade = E:
##             :       :       :...revol_util > 79: Fully Paid (19/2)
##             :       :       :   revol_util <= 79: [S16]
##             :       :       grade = B:
##             :       :       :...verification_status = Source Verified:
##             :       :       :   :...mort_acc <= 2: Fully Paid (67/30)
##             :       :       :   :   mort_acc > 2: Charged Off (16/3)
##             :       :       :   verification_status = Verified:
##             :       :       :   :...int_rate <= 10.16: Charged Off (9)
##             :       :       :   :   int_rate > 10.16:
##             :       :       :   :   :...annual_inc <= 26500: Charged Off (10)
##             :       :       :   :       annual_inc > 26500: [S17]
##             :       :       :   verification_status = Not Verified:
##             :       :       :   :...last_fico_range_low <= 620: Fully Paid (56/24)
##             :       :       :       last_fico_range_low > 620:
##             :       :       :       :...last_fico_range_low <= 625:
##             :       :       :           :...num_sats <= 22: Charged Off (36/5)
##             :       :       :           :   num_sats > 22: Fully Paid (2)
##             :       :       :           last_fico_range_low > 625: [S18]
##             :       :       grade = D:
##             :       :       :...last_fico_range_low > 630: Fully Paid (56/21)
##             :       :           last_fico_range_low <= 630: [S19]
##             :       funded_amnt > 8525:
##             :       :...annual_inc <= 63731: Charged Off (1226/327)
##             :           annual_inc > 63731:
##             :           :...funded_amnt > 14825: Charged Off (778/243)
##             :               funded_amnt <= 14825:
##             :               :...grade in {E,F}: Charged Off (11/4)
```

```
##             :                   grade = G: Fully Paid (1)
##             :                   grade = D:
##             :                   :...int_rate <= 18.85: Charged Off (42/10)
##             :                   :   int_rate > 18.85: Fully Paid (16/6)
##             :                   grade = A:
##             :                   :...acc_open_past_24mths <= 0: Fully Paid (4)
##             :                   :   acc_open_past_24mths > 0: [S20]
##             :                   grade = B:
##             :                   :...num_bc_tl <= 7: [S21]
##             :                   :   num_bc_tl > 7: [S22]
##             :                   grade = C: [S23]
##             last_fico_range_low > 635:
##             :...total_acc > 21:
##                 :...dti > 20.84:
##                 :   :...funded_amnt <= 5225: [S24]
##                 :   :   funded_amnt > 5225:
##                 :   :   :...grade in {A,B,E,G}: Charged Off (266/91)
##                 :   :       grade = F: Fully Paid (3/1)
##                 :   :       grade = D:
##                 :   :       :...tot_hi_cred_lim <= 351044: Charged Off (97/34)
##                 :   :       :   tot_hi_cred_lim > 351044: Fully Paid (10/1)
##                 :   :       grade = C:
##                 :   :       :...num_bc_tl > 16: Charged Off (29/4)
##                 :   :           num_bc_tl <= 16:
##                 :   :           :...dti <= 21.32: Charged Off (11)
##                 :   :               dti > 21.32: [S25]
##                 :   dti <= 20.84:
##                 :   :...open_acc > 20: Charged Off (76/28)
##                 :       open_acc <= 20:
##                 :       :...fico_range_low <= 680:
##                 :           :...last_fico_range_low > 640: Fully Paid (257/86)
##                 :           :   last_fico_range_low <= 640: [S26]
##                 :           fico_range_low > 680:
##                 :           :...int_rate > 12.12:
##                 :               :...grade in {A,C,E,F,
##                 :               :   :         G}: Charged Off (103/48)
##                 :               :   grade = D:
##                 :               :   :...dti <= 19.99: Charged Off (33/6)
##                 :               :   :   dti > 19.99: Fully Paid (5)
##                 :               :   grade = B: [S27]
##                 :               int_rate <= 12.12:
##                 :               :...num_bc_tl > 8:
##                 :                   :...total_acc <= 40: Fully Paid (121/49)
##                 :                   :   total_acc > 40: Charged Off (23/6)
##                 :                   num_bc_tl <= 8:
##                 :                   :...open_acc <= 13: Fully Paid (32/2)
##                 :                       open_acc > 13: [S28]
##                 total_acc <= 21:
##                 :...last_fico_range_low > 645:
##                     :...dti <= 21.61:
##                     :   :...fico_range_low <= 700:
##                     :   :   :...num_bc_tl <= 9: Fully Paid (185/28)
##                     :   :   :   num_bc_tl > 9: [S29]
##                     :   :   fico_range_low > 700:
```

```
##                       :   :   :...grade = A: Charged Off (9/4)
##                       :   :       grade in {D,E,F,G}: Fully Paid (2)
##                       :   :       grade = B:
##                       :   :       :...int_rate <= 9.99: Fully Paid (8)
##                       :   :       :   int_rate > 9.99: Charged Off (15/5)
##                       :   :       grade = C:
##                       :   :       :...bc_open_to_buy <= 5251: Fully Paid (3)
##                       :   :           bc_open_to_buy > 5251: Charged Off (4)
##                       :   dti > 21.61:
##                       :   :...open_acc > 13: Charged Off (7)
##                       :       open_acc <= 13:
##                       :       :...grade in {A,E,F,G}: Fully Paid (11/3)
##                       :           grade = D: Charged Off (16/6)
##                       :           grade = C:
##                       :           :...annual_inc <= 48000: Charged Off (16/6)
##                       :           :   annual_inc > 48000: Fully Paid (6)
##                       :           grade = B:
##                       :           :...total_acc <= 18: Fully Paid (16/3)
##                       :               total_acc > 18: [S30]
##                      last_fico_range_low <= 645:
##                      :...num_bc_sats > 7:
##                          :...num_bc_sats <= 11: Fully Paid (28/3)
##                          :   num_bc_sats > 11: Charged Off (2)
##                          num_bc_sats <= 7:
##                          :...funded_amnt <= 3925: Fully Paid (54/11)
##                              funded_amnt > 3925:
##                              :...grade in {C,G}: Fully Paid (142/59)
##                                  grade = A:
##                                  :...total_acc <= 12: Fully Paid (5)
##                                  :   total_acc > 12: Charged Off (41/18)
##                                  grade = E: [S31]
##                                  grade = F: [S32]
##                                  grade = B: [S33]
##                                  grade = D:
##                                  :...fico_range_low > 705: Charged Off (6)
##                                      fico_range_low <= 705:
##                                      :...installment > 705.59: Charged Off (12/1)
##                                          installment <= 705.59: [S34]
##
## SubTree [S1]
##
## total_rev_hi_lim <= 52219: Charged Off (24/2)
## total_rev_hi_lim > 52219: Fully Paid (2)
##
## SubTree [S2]
##
## acc_open_past_24mths <= 2: Charged Off (7)
## acc_open_past_24mths > 2: Fully Paid (8/3)
##
## SubTree [S3]
##
## mo_sin_old_rev_tl_op <= 128: Fully Paid (2)
## mo_sin_old_rev_tl_op > 128: Charged Off (12)
##
```

```
## SubTree [S4]
##
## total_bal_ex_mort <= 23879: Fully Paid (5/1)
## total_bal_ex_mort > 23879: Charged Off (5)
##
## SubTree [S5]
##
## bc_open_to_buy <= 788: Charged Off (2)
## bc_open_to_buy > 788: Fully Paid (4)
##
## SubTree [S6]
##
## verification_status = Verified: Fully Paid (31/6)
## verification_status = Not Verified:
## :...num_bc_tl <= 14: Fully Paid (72/15)
## :   num_bc_tl > 14:
## :   :...mo_sin_old_rev_tl_op <= 275: Charged Off (5)
## :       mo_sin_old_rev_tl_op > 275: Fully Paid (2)
## verification_status = Source Verified:
## :...grade in {A,E,F,G}: Fully Paid (16/5)
##     grade in {C,D}: Charged Off (8/2)
##     grade = B:
##     :...total_acc <= 21: Fully Paid (4)
##         total_acc > 21: Charged Off (2)
##
## SubTree [S7]
##
## acc_open_past_24mths <= 5: Fully Paid (6/1)
## acc_open_past_24mths > 5: Charged Off (6)
##
## SubTree [S8]
##
## num_bc_sats <= 4: Fully Paid (11/2)
## num_bc_sats > 4: Charged Off (3)
##
## SubTree [S9]
##
## last_fico_range_low > 655: Fully Paid (30/9)
## last_fico_range_low <= 655:
## :...fico_range_low <= 690: Fully Paid (5)
##     fico_range_low > 690: Charged Off (16/4)
##
## SubTree [S10]
##
## funded_amnt <= 4500: Fully Paid (4/1)
## funded_amnt > 4500: Charged Off (9)
##
## SubTree [S11]
##
## verification_status = Source Verified: Fully Paid (32/14)
## verification_status = Not Verified:
## :...open_acc <= 10: Fully Paid (6)
## :   open_acc > 10: Charged Off (25/11)
## verification_status = Verified:
```

```
## :...annual_inc <= 36400: Charged Off (11)
##     annual_inc > 36400:
##     :...num_bc_sats <= 8: Charged Off (44/16)
##         num_bc_sats > 8: Fully Paid (12/2)
##
## SubTree [S12]
##
## annual_inc <= 81500: Charged Off (6)
## annual_inc > 81500: Fully Paid (8/1)
##
## SubTree [S13]
##
## mort_acc <= 1: Fully Paid (9)
## mort_acc > 1: Charged Off (6/1)
##
## SubTree [S14]
##
## verification_status = Not Verified:
## :...num_bc_sats <= 3: Charged Off (4)
## :   num_bc_sats > 3: Fully Paid (2)
## verification_status in {Source Verified,Verified}:
## :...num_bc_sats <= 6: Fully Paid (28/2)
##     num_bc_sats > 6:
##     :...num_bc_sats <= 10: Charged Off (4)
##         num_bc_sats > 10: Fully Paid (2)
##
## SubTree [S15]
##
## verification_status in {Not Verified,Verified}: Fully Paid (47/22)
## verification_status = Source Verified:
## :...open_acc <= 4: Fully Paid (2)
##     open_acc > 4:
##     :...total_rev_hi_lim <= 12400: Charged Off (11)
##         total_rev_hi_lim > 12400: Fully Paid (2)
##
## SubTree [S16]
##
## verification_status = Source Verified: Charged Off (24/4)
## verification_status = Not Verified:
## :...fico_range_low <= 695: Charged Off (21/6)
## :   fico_range_low > 695: Fully Paid (2)
## verification_status = Verified:
## :...bc_open_to_buy <= 190: Fully Paid (5)
##     bc_open_to_buy > 190:
##     :...bc_open_to_buy <= 6151: Charged Off (21/3)
##         bc_open_to_buy > 6151: Fully Paid (2)
##
## SubTree [S17]
##
## fico_range_low <= 665: Fully Paid (9/1)
## fico_range_low > 665: Charged Off (53/23)
##
## SubTree [S18]
##
```

```
## last_fico_range_low > 630: Charged Off (35/14)
## last_fico_range_low <= 630:
## :...total_bal_ex_mort > 19066: Fully Paid (18/2)
##     total_bal_ex_mort <= 19066:
##     :...dti <= 4.29: Fully Paid (2)
##         dti > 4.29: Charged Off (8)
##
## SubTree [S19]
##
## verification_status = Not Verified: Charged Off (78/30)
## verification_status = Source Verified:
## :...revol_util > 43.2: Charged Off (48/18)
## :   revol_util <= 43.2:
## :   :...total_acc <= 43: Fully Paid (14/1)
## :       total_acc > 43: Charged Off (2)
## verification_status = Verified:
## :...tot_hi_cred_lim <= 19580: Fully Paid (5)
##     tot_hi_cred_lim > 19580:
##     :...tot_hi_cred_lim > 320152: Fully Paid (4)
##         tot_hi_cred_lim <= 320152:
##         :...bc_open_to_buy > 319: Charged Off (27/2)
##             bc_open_to_buy <= 319:
##             :...annual_inc <= 28000: Fully Paid (5)
##                 annual_inc > 28000:
##                 :...num_bc_sats <= 4: Charged Off (11/1)
##                     num_bc_sats > 4: Fully Paid (2)
##
## SubTree [S20]
##
## verification_status = Not Verified: Charged Off (21/5)
## verification_status = Source Verified:
## :...bc_open_to_buy <= 3467: Fully Paid (4)
## :   bc_open_to_buy > 3467: Charged Off (14/5)
## verification_status = Verified:
## :...mort_acc <= 4: Fully Paid (7/1)
##     mort_acc > 4: Charged Off (2)
##
## SubTree [S21]
##
## mo_sin_old_rev_tl_op <= 90: Charged Off (6)
## mo_sin_old_rev_tl_op > 90: Fully Paid (45/9)
##
## SubTree [S22]
##
## acc_open_past_24mths <= 2: Charged Off (14)
## acc_open_past_24mths > 2:
## :...last_fico_range_low <= 625: Charged Off (43/14)
##     last_fico_range_low > 625: Fully Paid (40/18)
##
## SubTree [S23]
##
## acc_open_past_24mths <= 1: Charged Off (7)
## acc_open_past_24mths > 1:
## :...dti <= 10.38:
```

```
##     :...total_acc > 19: Fully Paid (13)
##     :   total_acc <= 19:
##     :   :...int_rate <= 14.33: Charged Off (3)
##     :       int_rate > 14.33: Fully Paid (2)
##     dti > 10.38:
##     :...mort_acc > 5:
##         :...last_fico_range_low <= 630: Fully Paid (11/2)
##         :   last_fico_range_low > 630: Charged Off (3)
##         mort_acc <= 5:
##         :...verification_status in {Not Verified,
##             :                       Source Verified}: Charged Off (56/19)
##             verification_status = Verified:
##             :...open_acc <= 11: Fully Paid (6/1)
##                 open_acc > 11:
##                 :...revol_util <= 32.3: Fully Paid (2)
##                     revol_util > 32.3: Charged Off (10)
##
## SubTree [S24]
##
## verification_status = Source Verified: Fully Paid (12/1)
## verification_status in {Not Verified,Verified}:
## :...revol_util <= 75.4:
##     :...acc_open_past_24mths <= 4: Charged Off (21/4)
##     :   acc_open_past_24mths > 4: Fully Paid (30/12)
##     revol_util > 75.4:
##     :...annual_inc > 40500: Fully Paid (14)
##         annual_inc <= 40500:
##         :...total_acc <= 28: Charged Off (3)
##             total_acc > 28: Fully Paid (2)
##
## SubTree [S25]
##
## last_fico_range_low > 640: Fully Paid (108/50)
## last_fico_range_low <= 640:
## :...annual_inc <= 106500: Charged Off (53/16)
##     annual_inc > 106500: Fully Paid (3)
##
## SubTree [S26]
##
## verification_status = Not Verified: Fully Paid (41/14)
## verification_status = Source Verified:
## :...bc_open_to_buy <= 4065: Charged Off (28/10)
## :   bc_open_to_buy > 4065: Fully Paid (11)
## verification_status = Verified:
## :...dti <= 17.33: Charged Off (29/8)
##     dti > 17.33: Fully Paid (14/3)
##
## SubTree [S27]
##
## total_bal_ex_mort > 65710: Fully Paid (8/1)
## total_bal_ex_mort <= 65710:
## :...mo_sin_old_rev_tl_op <= 95: Fully Paid (3)
##     mo_sin_old_rev_tl_op > 95: Charged Off (34/6)
##
```

```
## SubTree [S28]
##
## verification_status in {Source Verified,Verified}: Fully Paid (7/1)
## verification_status = Not Verified:
## :...total_acc <= 31: Charged Off (6)
##     total_acc > 31: Fully Paid (2)
##
## SubTree [S29]
##
## acc_open_past_24mths <= 4: Fully Paid (17/4)
## acc_open_past_24mths > 4: Charged Off (9/2)
##
## SubTree [S30]
##
## verification_status = Not Verified: Fully Paid (4/1)
## verification_status in {Source Verified,Verified}: Charged Off (6)
##
## SubTree [S31]
##
## verification_status in {Not Verified,Verified}: Charged Off (21/8)
## verification_status = Source Verified: Fully Paid (10/2)
##
## SubTree [S32]
##
## verification_status = Not Verified: Charged Off (3)
## verification_status in {Source Verified,Verified}: Fully Paid (8/1)
##
## SubTree [S33]
##
## verification_status = Source Verified: Fully Paid (46/17)
## verification_status = Verified:
## :...acc_open_past_24mths <= 8: Fully Paid (47/14)
## :   acc_open_past_24mths > 8: Charged Off (3)
## verification_status = Not Verified:
## :...int_rate <= 9.71: Charged Off (7)
##     int_rate > 9.71:
##     :...last_fico_range_low > 640: Fully Paid (40/18)
##         last_fico_range_low <= 640:
##         :...num_bc_sats <= 5: Charged Off (41/17)
##             num_bc_sats > 5: Fully Paid (6)
##
## SubTree [S34]
##
## verification_status = Not Verified: Fully Paid (36/13)
## verification_status = Source Verified:
## :...last_fico_range_low > 640: Fully Paid (13/3)
## :   last_fico_range_low <= 640:
## :   :...mort_acc > 0: Charged Off (4)
## :       mort_acc <= 0:
## :       :...annual_inc <= 42520: Charged Off (5/1)
## :           annual_inc > 42520: Fully Paid (7)
## verification_status = Verified:
## :...num_bc_tl > 5: Charged Off (14/1)
##     num_bc_tl <= 5:
```

```
##     :...mo_sin_old_rev_tl_op > 335: Charged Off (3)
##         mo_sin_old_rev_tl_op <= 335:
##         :...acc_open_past_24mths <= 6: Fully Paid (13/1)
##             acc_open_past_24mths > 6: Charged Off (2)
##
##
## Evaluation on training data (73684 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##     307 9560(13.0%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##   33841  3002    (a): class Charged Off
##    6558 30283    (b): class Fully Paid
##
##
##  Attribute usage:
##
##  100.00% last_fico_range_low
##   25.98% term
##   14.61% total_acc
##    7.53% funded_amnt
##    7.29% grade
##    7.21% fico_range_low
##    4.37% verification_status
##    4.30% num_bc_tl
##    4.09% annual_inc
##    3.25% open_acc
##    3.05% dti
##    2.95% num_bc_sats
##    2.62% mo_sin_old_rev_tl_op
##    1.47% acc_open_past_24mths
##    1.21% int_rate
##    1.01% total_rev_hi_lim
##    0.86% mort_acc
##    0.53% tot_cur_bal
##    0.41% tot_hi_cred_lim
##    0.37% revol_util
##    0.34% bc_open_to_buy
##    0.21% total_bal_ex_mort
##    0.15% installment
##    0.05% num_sats
##    0.01% revol_bal
##
##
## Time: 1.7 secs
```

```
fittedc50train <- predict(modelc50, newdata = train[,-1])
```

```r
print(paste('Accuracy for train:', mean(fittedc50train == train$loan_status)))
```

```
## [1] "Accuracy for train: 0.870256772162206"
```

```r
# test

fittedc50test <- predict(modelc50, newdata = validation[,-1])

print(paste('Accuracy for test:', mean(fittedc50test == validation$loan_status)))
```

```
## [1] "Accuracy for test: 0.853554270824851"
```

```r
fittedc50test <- predict(modelc50, newdata = test[,-1])

print(paste('Accuracy for test:', mean(fittedc50test == test$loan_status)))
```

```
## [1] "Accuracy for test: 0.851471845608892"
```

```r
CrossTable(x = test$loan_status, y = fittedc50test,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  32748
##
##
##                  | fittedc50test
## test$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------|-------------|-------------|-------------|
##      Charged Off |       14747 |        1601 |       16348 |
##                  |       0.902 |       0.098 |       0.499 |
##                  |       0.819 |       0.109 |             |
##                  |       0.450 |       0.049 |             |
## -----------------|-------------|-------------|-------------|
##       Fully Paid |        3263 |       13137 |       16400 |
##                  |       0.199 |       0.801 |       0.501 |
##                  |       0.181 |       0.891 |             |
##                  |       0.100 |       0.401 |             |
## -----------------|-------------|-------------|-------------|
##     Column Total |       18010 |       14738 |       32748 |
##                  |       0.550 |       0.450 |             |
## -----------------|-------------|-------------|-------------|
##
##
```

```r
mean(test$loan_status == fittedc50test)
```

```
## [1] 0.8514718
```

C5.0 could be developed by boosting. I excluded summary of the new model because of ease of readability.

```
## Boosting the accuracy of decision trees
# boosted decision tree with 10 trials

modelc50boosted <- C5.0(loan_status~., train, trials = 30)

#modelc50boosted
#(modelc50boosted)
```

```
fittedc50trainboosted <- predict(modelc50boosted, newdata = train[,-1])

print(paste('Accuracy for boosted train data:', mean(fittedc50trainboosted == train$loan_status)))
```

```
## [1] "Accuracy for boosted train data: 0.907334020954346"
```

```
fittedc50validationboosted <- predict(modelc50boosted, newdata = validation[,-1])

print(paste('Accuracy for boosted test data:', mean(fittedc50validationboosted == validation$loan_status
```

```
## [1] "Accuracy for boosted test data: 0.856607768097061"
```

```
# test

fittedc50testboosted <- predict(modelc50boosted, newdata = test[,-1])

print(paste('Accuracy for boosted test data:', mean(fittedc50testboosted == test$loan_status)))
```

```
## [1] "Accuracy for boosted test data: 0.853914742885062"
```

```
CrossTable(x = test$loan_status, y = fittedc50testboosted,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  32748
##
##
##                  | fittedc50testboosted
## test$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------|-------------|-------------|-------------|
##      Charged Off |       14665 |        1683 |       16348 |
##                  |       0.897 |       0.103 |       0.499 |
##                  |       0.825 |       0.112 |             |
##                  |       0.448 |       0.051 |             |
## -----------------|-------------|-------------|-------------|
##       Fully Paid |        3101 |       13299 |       16400 |
##                  |       0.189 |       0.811 |       0.501 |
```

```
##                  |       0.175 |       0.888 |             |
##                  |       0.095 |       0.406 |             |
## -----------------|-------------|-------------|-------------|
##     Column Total |       17766 |       14982 |       32748 |
##                  |       0.543 |       0.457 |             |
## -----------------|-------------|-------------|-------------|
##
##
```

```
mean(test$loan_status == fittedc50testboosted)
```

```
## [1] 0.8539147
```

Random Forest via ranger

Accuracy is perfect for the train dataset. Accuracy is 84.64% for test dataset again by ranger package.

```
tic()
rfranger <- ranger(loan_status ~ ., data = train, num.threads = 12, num.trees = 20, max.depth = 10)

rfranger
```

```
## Ranger result
##
## Call:
##  ranger(loan_status ~ ., data = train, num.threads = 12, num.trees = 20,      max.depth = 10)
##
## Type:                             Classification
## Number of trees:                  20
## Sample size:                      73684
## Number of independent variables:  25
## Mtry:                             5
## Target node size:                 1
## Variable importance mode:         none
## Splitrule:                        gini
## OOB prediction error:             14.72 %
```

```
rfranger$confusion.matrix
```

```
##              predicted
## true          Charged Off Fully Paid  <NA>
##   Charged Off       33155       3685     3
##   Fully Paid         7163      29677     1
```

```
summary(rfranger)
```

```
##                           Length Class         Mode
## predictions               73684  factor        numeric
## num.trees                     1  -none-        numeric
## num.independent.variables     1  -none-        numeric
## mtry                          1  -none-        numeric
## min.node.size                 1  -none-        numeric
## prediction.error              1  -none-        numeric
## forest                        9  ranger.forest list
## confusion.matrix              6  table         numeric
## splitrule                     1  -none-        character
## treetype                      1  -none-        character
```

```
## call                            6  -none-        call
## importance.mode                 1  -none-        character
## num.samples                     1  -none-        numeric
## replace                         1  -none-        logical
```

```
rf_predrangertrain <- predict(rfranger, train)

confusionMatrix(data=rf_predrangertrain$predictions, train$loan_status)
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    Charged Off Fully Paid
##   Charged Off       33806       6772
##   Fully Paid         3037      30069
##
##                Accuracy : 0.8669
##                  95% CI : (0.8644, 0.8693)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7338
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9176
##             Specificity : 0.8162
##          Pos Pred Value : 0.8331
##          Neg Pred Value : 0.9083
##              Prevalence : 0.5000
##          Detection Rate : 0.4588
##    Detection Prevalence : 0.5507
##       Balanced Accuracy : 0.8669
##
##        'Positive' Class : Charged Off
##
```

```
rf_predrangerval <- predict(rfranger, validation)

confusionMatrix(data=rf_predrangerval$predictions, validation$loan_status)
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    Charged Off Fully Paid
##   Charged Off       11176       2370
##   Fully Paid         1130       9886
##
##                Accuracy : 0.8575
##                  95% CI : (0.8531, 0.8619)
##     No Information Rate : 0.501
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7149
##
```

```
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9082
##             Specificity : 0.8066
##          Pos Pred Value : 0.8250
##          Neg Pred Value : 0.8974
##              Prevalence : 0.5010
##          Detection Rate : 0.4550
##    Detection Prevalence : 0.5515
##       Balanced Accuracy : 0.8574
##
##        'Positive' Class : Charged Off
##
```

```
toc()
```

```
## 2 sec elapsed
```

```
rf_predrangertest <- predict(rfranger, test)

confusionMatrix(data=rf_predrangertest$predictions, test$loan_status)
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    Charged Off Fully Paid
##   Charged Off       14861       3266
##   Fully Paid         1487      13134
##
##                Accuracy : 0.8549
##                  95% CI : (0.851, 0.8587)
##     No Information Rate : 0.5008
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7098
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9090
##             Specificity : 0.8009
##          Pos Pred Value : 0.8198
##          Neg Pred Value : 0.8983
##              Prevalence : 0.4992
##          Detection Rate : 0.4538
##    Detection Prevalence : 0.5535
##       Balanced Accuracy : 0.8549
##
##        'Positive' Class : Charged Off
##
```

```
CrossTable(x = test$loan_status, y = rf_predrangertest$predictions,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
```

```
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  32748
##
##
##                  | rf_predrangertest$predictions
## test$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------|-------------|-------------|-------------|
##      Charged Off |       14861 |        1487 |       16348 |
##                  |       0.909 |       0.091 |       0.499 |
##                  |       0.820 |       0.102 |             |
##                  |       0.454 |       0.045 |             |
## -----------------|-------------|-------------|-------------|
##       Fully Paid |        3266 |       13134 |       16400 |
##                  |       0.199 |       0.801 |       0.501 |
##                  |       0.180 |       0.898 |             |
##                  |       0.100 |       0.401 |             |
## -----------------|-------------|-------------|-------------|
##     Column Total |       18127 |       14621 |       32748 |
##                  |       0.554 |       0.446 |             |
## -----------------|-------------|-------------|-------------|
##
##
```

```
mean(test$loan_status == rf_predrangertest$predictions)
```

```
## [1] 0.8548614
```

Model Evaluation: Lending Club 2015 Data

Once you have decided on the best model, refit it using all of the 2012-2014 data and then use your model to classify all of the 2015 data. Check the accuracy of your predictions.

2015 Lending Club data includes higher proportions of charged-off category due to 60 months credits mostly have not resulted for fully-paid credits.

```
LendingClub_2015 <- LendingClub %>%
  filter(year == 2015) %>%
  select(loan_status, funded_amnt, annual_inc, term, installment, int_rate, grade, dti,
         verification_status, fico_range_low, total_acc, tot_cur_bal, acc_open_past_24mths, num_bc_sats
         open_acc, revol_bal, revol_util, last_fico_range_low, total_rev_hi_lim,
         bc_open_to_buy, mo_sin_old_rev_tl_op, mort_acc, num_bc_sats, num_bc_tl, num_sats,
         tot_hi_cred_lim, total_bal_ex_mort)

LendingClub_2015 <- LendingClub_2015 %>%
  filter(loan_status == "Charged Off"  | loan_status == "Fully Paid") %>% na.omit()

LendingClub_2015$loan_status <- factor(LendingClub_2015$loan_status)
levels(LendingClub_2015$loan_status)
```

```
## [1] "Charged Off" "Fully Paid"
```

```
addmargins(table(LendingClub_2015$loan_status, LendingClub_2015$term))
```

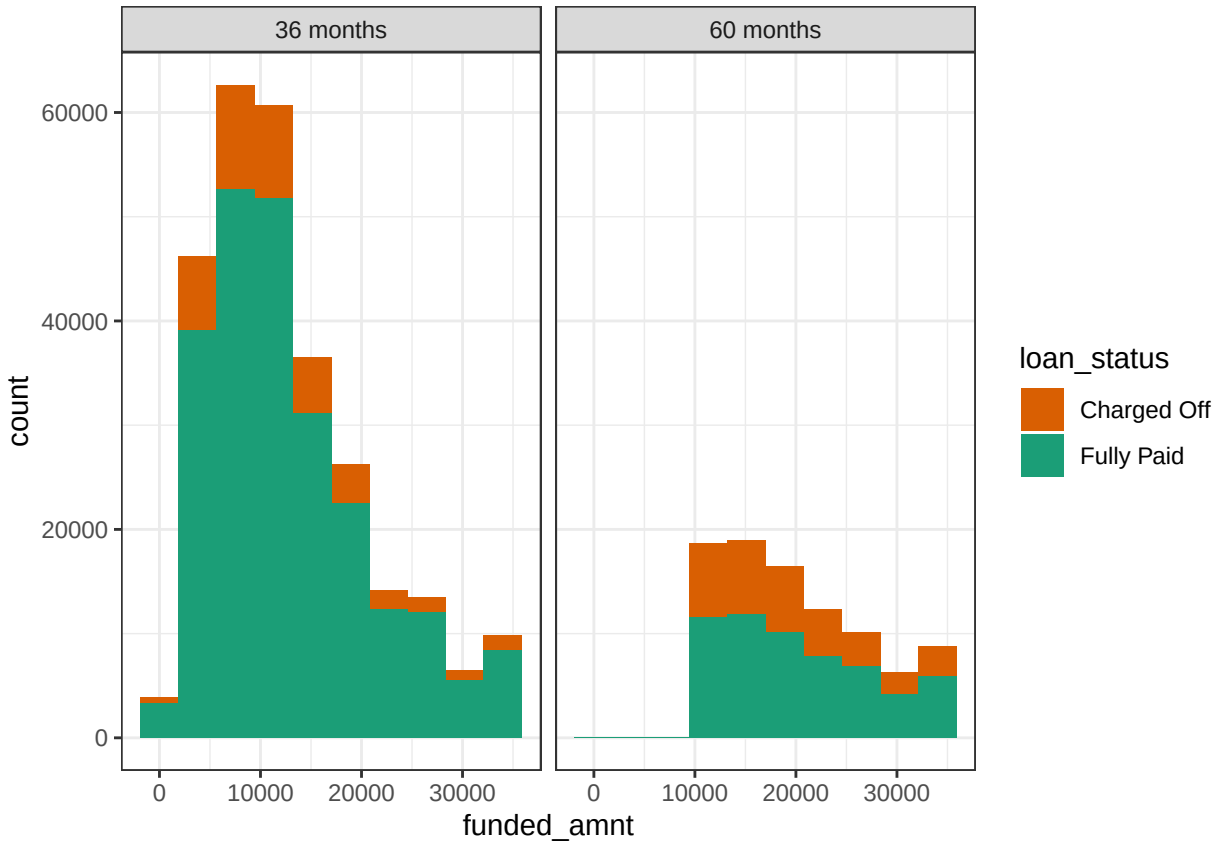
```
##
##                36 months 60 months    Sum
##   Charged Off      41649     33362  75011
##   Fully Paid      238524     58367 296891
##   Sum             280173     91729 371902
```

```
addmargins(prop.table(table(LendingClub_2015$loan_status, LendingClub_2015$term)))
```

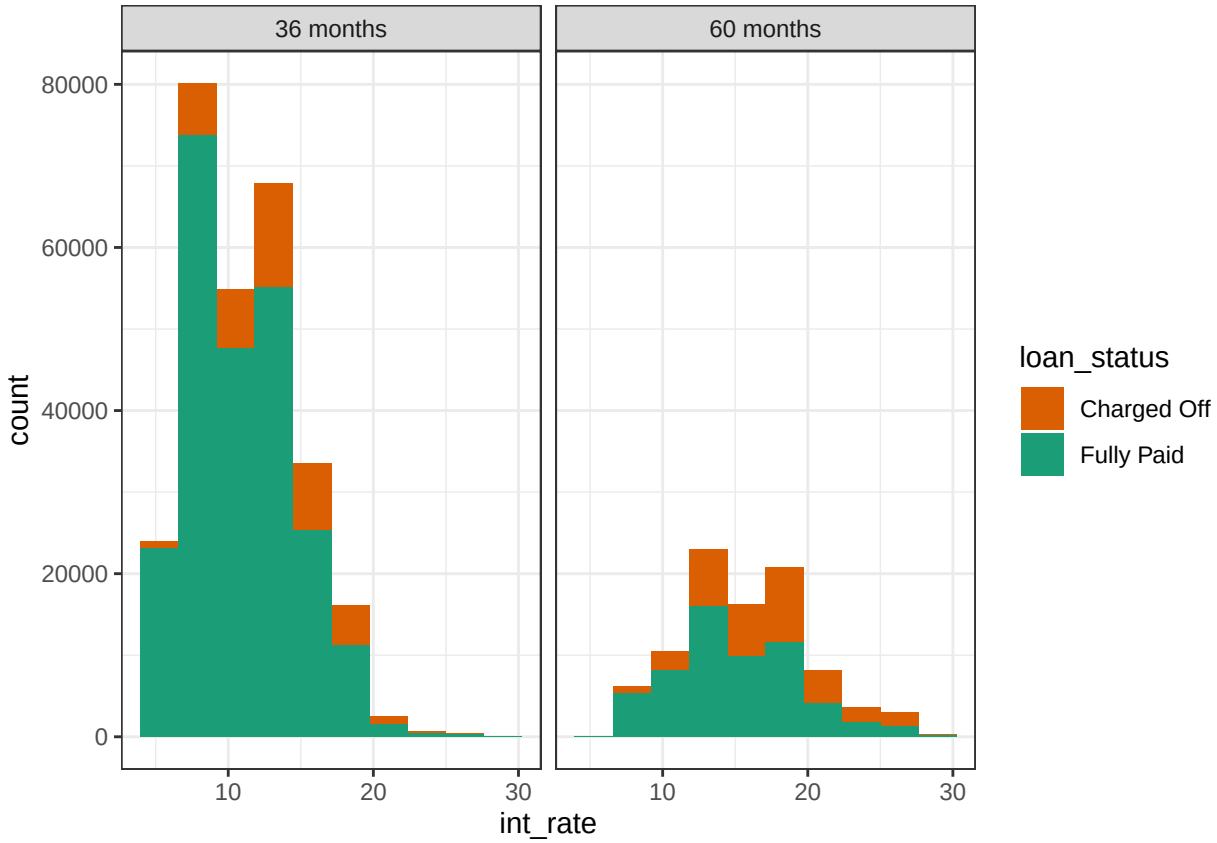
```
##
##                 36 months  60 months        Sum
##   Charged Off 0.11198918 0.08970643 0.20169561
##   Fully Paid  0.64136251 0.15694188 0.79830439
##   Sum         0.75335169 0.24664831 1.00000000
```

```
options(repr.plot.width = 1, repr.plot.height = 0.5)

LendingClub_2015 %>%
  ggplot(aes(funded_amnt, fill = loan_status)) +
  geom_histogram(bins = 10) +
  scale_fill_brewer(palette = "Dark2", direction = -1) +
  facet_wrap(~term) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```



```
LendingClub_2015 %>%
  ggplot(aes(int_rate, fill = loan_status)) +
  geom_histogram(bins = 10) +
  scale_fill_brewer(palette = "Dark2", direction = -1) +
  facet_wrap(~term) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```




```
dim(LendingClub_2015)
```

```
## [1] 371902     26
```

```
LendingClub_2015 %>%
  group_by(loan_status) %>%
  summarise(n = n()) %>%
  mutate(freq = n/sum(n))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 3
##   loan_status      n  freq
##   <fct>        <int> <dbl>
## 1 Charged Off  75011 0.202
## 2 Fully Paid  296891 0.798
```

```
pred <- predict(log_reg, newdata = LendingClub_2015, type = "response")

log_reg_pred <- ifelse(pred > 0.5, "Fully Paid", "Charged Off")
```

```
CrossTable(x = LendingClub_2015$loan_status, y = log_reg_pred,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  371902
##
##
##                              | log_reg_pred
## LendingClub_2015$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------------------|-------------|-------------|-------------|
##                  Charged Off |       67965 |        7046 |       75011 |
##                              |       0.906 |       0.094 |       0.202 |
##                              |       0.659 |       0.026 |             |
##                              |       0.183 |       0.019 |             |
## -----------------------------|-------------|-------------|-------------|
##                   Fully Paid |       35111 |      261780 |      296891 |
##                              |       0.118 |       0.882 |       0.798 |
##                              |       0.341 |       0.974 |             |
##                              |       0.094 |       0.704 |             |
## -----------------------------|-------------|-------------|-------------|
##                 Column Total |      103076 |      268826 |      371902 |
##                              |       0.277 |       0.723 |             |
## -----------------------------|-------------|-------------|-------------|
##
##
```

```
mean(LendingClub_2015$loan_status == log_reg_pred)
```

```
## [1] 0.8866449
```

```
LendingClub_2015_knn <- LendingClub_2015 %>% mutate_if(is.factor, as.numeric)

# Normalization
LendingClub_2015_knn_n  <- as.data.frame(lapply(LendingClub_2015_knn[2:26], normalize))

pred_knn <- knn(train_knn_n, LendingClub_2015_knn_n, cl= train_knn$loan_status, k=20)

# Evaluating model performance
CrossTable(x = LendingClub_2015_knn$loan_status, y = pred_knn,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
```

```
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  371902
##
##
##                                  | pred_knn
## LendingClub_2015_knn$loan_status |         1 |         2 | Row Total |
## ---------------------------------|-----------|-----------|-----------|
##                                1 |     57850 |     17161 |     75011 |
##                                  |     0.771 |     0.229 |     0.202 |
##                                  |     0.626 |     0.061 |           |
##                                  |     0.156 |     0.046 |           |
## ---------------------------------|-----------|-----------|-----------|
##                                2 |     34568 |    262323 |    296891 |
##                                  |     0.116 |     0.884 |     0.798 |
##                                  |     0.374 |     0.939 |           |
##                                  |     0.093 |     0.705 |           |
## ---------------------------------|-----------|-----------|-----------|
##                     Column Total |     92418 |    279484 |    371902 |
##                                  |     0.249 |     0.751 |           |
## ---------------------------------|-----------|-----------|-----------|
##
##
```

```
mean(LendingClub_2015_knn$loan_status == pred_knn)
```

```
## [1] 0.8609069
```

```
fittedc50test <- predict(modelc50, newdata = LendingClub_2015[,-1])

print(paste('Accuracy for test:', mean(fittedc50test == LendingClub_2015$loan_status)))
```

```
## [1] "Accuracy for test: 0.86039064054509"
```

```
CrossTable(x = LendingClub_2015$loan_status, y = fittedc50test,
           prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  371902
##
```

```
##
##                              | fittedc50test
## LendingClub_2015$loan_status | Charged Off |  Fully Paid |   Row Total |
## -----------------------------|-------------|-------------|-------------|
##                  Charged Off |       70409 |        4602 |       75011 |
##                              |       0.939 |       0.061 |       0.202 |
##                              |       0.598 |       0.018 |             |
##                              |       0.189 |       0.012 |             |
## -----------------------------|-------------|-------------|-------------|
##                   Fully Paid |       47319 |      249572 |      296891 |
##                              |       0.159 |       0.841 |       0.798 |
##                              |       0.402 |       0.982 |             |
##                              |       0.127 |       0.671 |             |
## -----------------------------|-------------|-------------|-------------|
##                 Column Total |      117728 |      254174 |      371902 |
##                              |       0.317 |       0.683 |             |
## -----------------------------|-------------|-------------|-------------|
##
##
```

```
mean(LendingClub_2015$loan_status == fittedc50test)
```

```
## [1] 0.8603906
```

```
rf_predrangertest <- predict(rfranger, LendingClub_2015)

confusionMatrix(data=rf_predrangertest$predictions, LendingClub_2015$loan_status)
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    Charged Off Fully Paid
##   Charged Off       70581      47051
##   Fully Paid         4430     249840
##
##                Accuracy : 0.8616
##                  95% CI : (0.8605, 0.8627)
##     No Information Rate : 0.7983
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6454
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9409
##             Specificity : 0.8415
##          Pos Pred Value : 0.6000
##          Neg Pred Value : 0.9826
##              Prevalence : 0.2017
##          Detection Rate : 0.1898
##    Detection Prevalence : 0.3163
##       Balanced Accuracy : 0.8912
##
##        'Positive' Class : Charged Off
##
```

Performance Measures

- Sensitivity = True Positives / (True Positives + False Negatives)
- Specificity = True Negatives / (True Negatives + False Positives)
- Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

Testing All Models on 2015 Data

Performances of all models for an imbalanced data can be seen below. Logistic Regression has better accuracy than others for 2015. However, k-nearest neighbors has the highest specificity and Random forest has the highest sensitivity.

Models	Sensitivity	Specificity	Accuracy
Logistic Regression	0.906	0.882	0.887
k-Nearest Neighbors	0.711	0.920	0.811
C5.0	0.939	0.841	0.860
Random Forest	0.944	0.835	0.855

Conclusion

Trade-off of this study is between the cost of default and paid credits.

My last decision is to use Random forest, the method with the highest sensitivity. Detecting bad credits correctly 94.4% instead 90.6% is more valuable than getting higher total accuracy by 2.2%.

If the loss and profit values are known in advance, then an integer programming approach could be implemented to make the final model selection easier.